Identifying Causes in Psychiatry

Lena Kästner
Ruhr-Universität Bochum, Institut für Philosophie II
mail@lenakaestner.de

Abstract

Explanations in psychiatry often integrate various factors relevant to psychopathology. Identifying genuine causes among them is theoretically and clinically important, but epistemically challenging. Woodward's interventionism appears to provide a promising tool to achieve this. However, it cannot be applied to psychiatry. I thus introduce *difference-making interventionism* (DMI), which detects relevance in general rather than causation. DMI mirrors the empirical reality of psychiatry even more closely than interventionism, but it needs to be supplied with additional heuristics to disambiguate between causes and other difference-makers. To achieve this, I suggest employing heuristics based on multiple experiments, temporal order and scientific domain.

Keywords

word count (incl. references and footnotes): 4674

Paper submitted to: Philosophy of Science Association

1. Causal Explanations in Psychiatry

Causal explanations are abundant in psychiatry. Common statements include, e.g., that traumatic life events cause depressions, that excess dopamine causes mania, that fear causes patients' heart rates to accelerate, and that a spiders crawling on an arachnophobic's arm caused her panic attack. Philosophers traditionally struggle with such causal claims linking processes in the physical domain (dopamine levels, heart rate, a spider's crawling on skin) to those in the realm of the mental (trauma, depression, fear, panic attack). But despite convincing theoretical arguments to the effect that the mental is causally pre-empted by the physical (e.g. Kim 1998), talk of causation is pervasive in the empirical reality of psychiatry.

One may suspect, of course, that causal explanations in psychiatry are merely an instance of "sloppy talk". But while in everyday clinical practice psychiatrists might sometimes use "cause" as shorthand for "difference-maker", psychiatry as a scientific discipline cannot afford to blur the distinction between genuine causes and other difference-makers (background conditions, parts of a whole, realizers, etc.). Building models of psychiatric diseases, disambiguating between alternative explanations, and selecting treatment options all may require understanding what genuine causes of a given condition are. Memory loss after a car crash, for instance, may result from posttraumatic stress disorder (PTSD) or gas poisoning that occurred during the accident. Clearly, the treatment will differ depending on whether poisoning or trauma caused memory loss. Distinguishing causes from other difference-makers in explanations of mental illness thus is not merely a philosophical apprentice piece but highly clinically relevant.

With his manipulability-based *interventionism*, Woodward (2003) proposed an account of causation and causal explanation that aligns nicely with experimental research practice. As such, it seems a promising candidate to overcome (philosophical) worries of causal exclusion and evaluate causal claims in psychiatry (Kendler and Campbell 2009, p. 886). However, interventionism is not up to the task: it is designed to assess causal relations *only* in contexts where no other dependence relations are present. But psychiatrists' causal claims often relate mental (memory loss, depression) and physical (car crash,

neurotransmitters) domains, which—problematically for the interventionist—are usually considered to stand in non-causal dependence relations. No matter whether we characterize cognitive processes as *based* on, *grounded* in, *implemented*, *realized*, *constituted*, or *underlain* by, or *supervening* on neurophysiological process in the brain: as soon as non-causal relations are at issue, interventionism cannot be applied.

This problem is a principled one; it shows up in many special sciences exhibiting a certain inter- or multi-level character. Many modern special sciences combine insights gained at different levels of investigation, from different scientific domains. The same principled questions we are dealing with when relating psychology and neuroscience in psychiatric explanations also arise when linking factors from, say, astronomy, and physics, molecular biology and population genetics, or anthropology and meteorology. But for the time being let us focus on psychiatry.

The current paper offers a systematic account of how to identify causes in psychiatry despite the presence of non-causal dependence relations.[1] I suggest a weakened version of interventionism, *difference-making interventionism* (DMI), that allows applying interventionist methodology despite the presence of non-causal dependence relations. DMI acknowledges that observed manipulability underdetermines the underlying dependence relation. But this loss in specificity is a virtue rather than a vice: Psychiatrists often employ interventions (e.g. in randomized control trials) *before* they know what kind of relation grounds the relevance of difference-makers. Likewise, recent attempts to model psychiatric disorders by means of networks (e.g. Borsboom 2017) incorporate a range of different factors (physiological, genetic, environmental, behavioral), regardless of the possible dependence relations between them. DMI captures and accommodates for this. Besides, it highlights that thinking about explanatory relevance in terms of causal relevance only is too narrow. Still, embracing DMI does neither mean causal relevance loses its special status, nor that we cannot disambiguate between different kinds of dependence relations *in principle*.

---

[1] This is not a paper on the metaphysics or semantics of causation. I am here primarily concerned with the epistemic question of how to identify causal relations in cases where other dependence relations are present as well, particularly in psychiatry.

Disambiguation between causal and non-causal dependencies might be achieved, I suggest, by drawing on resources other than interventions: We may not only combine evidence from *many experiments* to infer systematic dependencies but can also supplement difference-making graphs with the *dimensions* of *time* and *domain*. A third dimension may accommodate for *multiple realizers*. Thus equipped, DMI preserves the manipulationist core of interventionism while relying on additional heuristics helps identify causes among difference-makers.

Section 2 introduces Woodward's interventionism and the problems with applying it to psychiatry. Section 3 introduces DMI along with some heuristics that may help us identify genuine causes. Section 4 concludes.

2. Woodward's Interventionism and its Application in Psychiatry

According to Woodward's (2003, 2008) *interventionist account of causation*, causal relations can be detected by difference-making: causes make differences to their effects. Causal explanations thus embody a "what-if-things-had-been-different conception of explanation" (Woodward 2003, p. 228). That is, they tell us what will (or would) happen under a range of different circumstances. Inspired by work on causal modeling (Spirtes et al. 1993, Pearl 2000), Woodward translates questions about causal relations into questions about relations between variables (representing properties or events) taking different values. Causal relations between variables can be represented in *directed acyclic graphs*. On the interventionist account, we can infer that X causes Y if and only if we can carry out an intervention on X with respect to Y. A manipulation of X qualifies as an *intervention I on X with respect to Y* if and only if I meets the following conditions: (i) I causes X, (ii) I overrides all other causes of X, (iii) any directed path from I to Y goes through X (i.e. there must not be a causal path, neither direct nor through other variables, from I to Y that does not go through X), and (iv) I is statistically independent of any variable Z that causes Y and that is on a directed path that does not go through X. That is to say, an intervention I

"breaks off" all other influences on X and manipulates X in such a way that changes in Y are *only* mediated through changes in C and *not in any other way*.[2]

Interventionism thus defined mirrors the manipulative character of experimental research practice and reflects certain well-known principles of experimental design (e.g. randomized control trials). It thus allows us to *directly infer* causal relations on the basis of observed manipulability in well-designed empirical studies. Prima facie then, assuming adequate experimental standards apply in psychiatry, we might think that the interventionist view "provides a single, clear empirical framework for the evaluation of all causal claims in psychiatry" (Kendler and Campbell 2009, p. 886; see also Campbell 2007, 2016, Rescorla 2017).

Consider the following toy examples: Tom was mentally healthy before his father died. But as part of his bereavement reaction Tom started grieving. While initial grief after bereavement is not a mental illness, Tom's bereavement experience was so severe and long-lasting that it developed into full-blown depression, although nothing else had changed about his life.[3] Given this picture, it makes sense for the psychiatrist to infer that Tom's grief (more precisely, his bereavement experience characterized by severe grief) caused his depression. Similarly, if Gina's heart rate was at 69bpm while she was relaxing on the sofa before it suddenly accelerated to 132bpm as she was afraid there was a burglar in the hallway (while nothing else changed) it makes sense to infer that Gina's fear of a burglar caused her heart rate to accelerate. And so on. We can picture this with the graphs are shown in figure 1.

---

[2] Woodward captures this in his definitions (M) and (IV). While (M) expresses that causation is grounded in manipulability relations, (IV) explicates the conditions under which a manipulation is an appropriate intervention, i.e. suitable to uncover causal relations. For the full definitions see Woodward (2003) p. 59 and p. 98, respectively.

[3] The precise relations between bereavement experiences (including grief, apathy, guilt, etc.) and depression are more complex than the current example suggests (see, e.g., Pies 2014, Wagner 2014 for discussions about dropping the bereavement exclusion in DSM V). For current purposes, however, the simplified scenario of Tom developing a depression some time after his bereavement experience, which was primarily characterized by severe grief, will do. For brevity, I shall talk about Tom's grief.

$$I \longrightarrow G \longrightarrow D$$

$$I \longrightarrow F \longrightarrow H$$

Figure 1: Two independent causal graphs illustrating Gina's and Tom's cases. Setting the value of G (representing Tom's grief) from 0 to 1 changes the value of D (representing his depressive state) from 0 to 1. Setting the value of F (representing Gina's fear) from 0 to 1 changes the value of H (representing her heart rate) from 69 to 132bpm. The interventions (I) are the death of Tom's father and Gina's hearing footsteps in the hallway.

This reasoning squares well with clinical evidence to the effect that grief can cause depressive episodes (e.g. Beck and Alford 2009) and that mental states such as fear affect an individual's heart rate (e.g. Cuthbert et. al. 2003). Likewise, there is evidence for manipulations in the mental domain, like cognitive behavioral therapy (CBT), to be efficacious in treating conditions such as anxiety and depression (e.g. Sofronoff, Attwood, and Hinton 2005, Butler et al. 2006). But should this lead to us conclude that interventionism allows us to establish notoriously difficult mental-to-mental[4] (in Tom's case) as well as mental-to-physical (in Gina's case) causal claims? No. The cases are not as unproblematic for the interventionist as they might seem.

To ensure we can sort actual causes from confounding factors and accidental correlates, interventionist analyses require that all of the considered variables in a causal graph must in principle be *independently manipulable* (see Woodward 2008, p. 209, Woodward 2015). For if this were not the case, we will run risk of violating (iii) and (iv). Thus, if we assume there is some sort of systematic (implementation, realization, supervenience, grounding, …) relation between mental and physical phenomena, applying interventionism is—despite its intuitive plausibility—simply *not licensed* (Eronen 2012, Raatikainen 2010, Shapiro 2010, Shapiro & Sober 2007, Kästner 2017).

Recently, a number of attempts have been made to save interventionism for scenarios with non-causal relations (for discussions in the context of mental causation see, e.g., Woodward 2008, 2015, Baumgartner 2010, 2013, Gebharter

---

[4] For the sake of the example suppose Tom's depression (at least his depressive mood) is a mental phenomenon.

2015, Hoffmann-Kolss 2014, Kästner 2017; an analogous debate in the context of constitutive mechanistic explanations is reviewed in Kästner and Andersen 2018). The proposed modifications typically advocate either splitting causal graphs or introducing exception-clauses for non-causal dependence relations. But even if these strategies were successful, neither is convincing in the case of psychiatry. First, psychiatrists typically aim at *integrated* explanations relating different (mental, neurophysiological, genetic, ...) factors. These factors may be relevant for different reasons: because they exert a causal influence, because they are a part of the implementational (realization, supervenience) base of a certain psychopathology, because they are background conditions, etc. Interventionism, by contrast, is designed to assess *causal relations only*. Second, the exact relations between different variables often remain subject to investigation and cannot be presupposed before scientists start testing for manipulability.

Current network models of mental disorders (e.g. Borsboom 2017, Borsboom, Cramer & Kalis forthcoming) illustrate this. These models are typically based on interventionist reasoning. However, they usually include concrete symptoms along with behavioral, cognitive, genetic, demographic, and environmental factors as variables. While some of these may in fact be related causally (e.g. grief causing depressive mood), for others that seems at least questionable. Is Tom's low serotonin, for instance, implementing or causing depression?[5] And does his socio-economic situation causally contribute or is it merely a background condition? Despite such questions, network models are powerful tools to figure out which factors are relevant to mental disorders. Applying interventionist reasoning to them helps uncover which factors influence one another, as well as what the developmental dynamics are. Moreover, the manipulationist strategy matches well with empirical research reality and provides our best currently available account of scientific (causal) explanation. Thus, it seems well worth trying to save interventionism for psychiatry.

---

[5] Talking about "low serotonin" as the substrate of depression is probably too simplistic; you might consider it a placeholder for whatever the neurophysiological substrate of depression according to your favorite account.

3. Difference-Making Interventionism for Psychiatry

To save interventionist reasoning for psychiatry, I introduce a weakened form of interventionism: *difference-making interventionism* (DMI). Rather than limiting our analysis to causal relations, DMI identifies a whole bundle of difference-making relations (among them, of course, causal relations). Thus, we can apply DMI in cases where non-causal dependence relations are present. To balance the resulting loss in specificity, we can employ additional heuristics allowing us to identify genuine causes.

3.1 The Bare Bones of DMI

To uncover dependence relations with DMI, we can keep using variables and directed graphs (now speaking of difference-making rather than causal paths) and proceed by the familiar interventionist method: DMI takes X to be a difference-maker for Y with respect to a given variable set V if and only if there is a possible (IV_{dm})-defined intervention on X with respect to Y that will change Y when all other variables Z_i in V are held fixed, except for those on a *difference-making path* from X to Y. (IV_{dm}) defines an intervention as follows: A manipulation I of X qualifies as an *intervention on X with respect to Y* if and only if it meets the following conditions: (i) I is a difference-maker for X, (ii) I overrides all other difference-makers influencing X, (iii) any directed path from I to Y goes through X (i.e. there must not be a difference-making path, neither direct nor through other variables, from I to Y that does not go through X), and (iv) I is statistically independent of any variable Z_i in V making a difference to Y and that is on a directed path that does not go through X.[6]

The advantages of DMI are that we no longer need to worry about restricting our variable set to independently manipulable variables or knowing among which variables non-causal dependencies obtain *before* we proceed to test for manipulability. This is empirically realistic as it reflects that (a) conceiving of explanatory relevance as causal relevance *only* is too narrow and (b) psychiatrists often employ interventionist reasoning (e.g. in randomized

---

[6] This is structurally analogous to Woodward's definitions (M) and (IV), just modified to no longer restrict our analysis to causal relations (cf. section 2).

control trials) *before* they know what kind of relation grounds the observed difference-making relation. The clinical efficacy of antidepressants, for instance, underdetermines why these drugs work. Do they work because they target the cause of disease or because they interfere with the pharmacological mechanism implementing certain pathologies? Likewise, does CBT help alleviate Tom's depression because it directly targets his mood or because it otherwise induces changes in, say, the low serotonin levels underlying his depression? DMI explicitly acknowledges this underdetermination.

However, the caveat is a significant loss in specificity. Once we adopt DMI, manipulability can no longer be used to directly infer causal relations (otherwise we would face an inflation of causal claims!); DMI underdetermines the underlying dependence relation. But, I propose, this is a virtue rather than a vice: integrating causal and other explanatory factors into a single model is a key feature of network explanations of mental disorders (e.g. Borsboom 2017; Borsboom, Cramer & Kalis forthcoming). Of course, we still need some way or other to identify causes among explanatorily relevant factors. To achieve this, I suggest, we may supplement interventions with other strategies.

3.2. Heuristic Inferences: Asymmetry and Multiple Experiments

It is a platitude about causation that causes are spatiotemporally distinct from their effects (e.g. Lewis 1970); causes precede their effects, and effects depend on their causes but not vice versa. Building on this knowledge, we gain at least two possible criteria to distinguish causation from other forms of dependence: time and asymmetry.[7] Let us first consider asymmetry.

Asymmetric manipulability is a first indication of but not by itself sufficient to infer causation. Take Gina's case: suppose we find that as the footsteps (I_1) induced Gina's fear (F) her heart rate (H) accelerated. But we can also get her heart rate to accelerate if we put her on a treadmill (I_2), which does not induce Gina's fear. This may lead us to infer that F causes H since we can intervene into F with respect to H but not vice versa. However, H may also be a supervenience

---

[7] Both strategies only work so long as we do not commit to simultaneous causation. Feedback loops can, however, be accommodated for in terms of repeated causal interactions between the same factors at different points in time (A causes B at t1 and B causes A at t2, ...) once we take into account temporal order and draw out feedback loops over time (see section 3.3).

base or realizer of F such that changes in F are necessarily accompanied by changes in H while *only some* changes in H will be accompanied by changes in F (and for the treadmill it was not).

We can thus derive the following heuristic for identifying causes *across multiple experiments* or repetitions: provided that repeated interventions on F with respect to H do affect H, we should consider F a genuine cause of H when a critical number of interventions (say, 1.000) into H with respect to F fails. If, by contrast, some of these interventions into H with respect to F actually affect F, it seems more plausible that F supervenes on / is realized by H.[8]

3.3. Adding Dimensions: Time, Domain and Multiple Realizers

Let us turn to time. Causation is typically considered *diachronic* (see fn 7) while non-causal dependence relations like realization, implementation, constitution, supervenience, part-whole, etc. are usually considered *synchronic* in nature. Thus, the *temporal profile* of variables changing their values in response to a given intervention may give us a clue as to whether or not variables are causally related. Notice, however, that inferences based on temporal order may be compromised by practical and methodological constraints. How quickly can my measurement technique detect changes? What are adequate timescales to consider (generations, hours, nanoseconds, …)? And how should I individuate variables to begin with?

While variable individuation is a notorious problem for the interventionist (and constraints will likely depend on the purpose of our analysis as well as the nature of the scenario) time often is already implicit in how we draw causal graphs: from left to right, starting with variables representing early occurrences of properties or events. Yet, what matters for interventionists merely is which variables are linked through edges, not how they are positioned. By adding an "arrow of time" into the picture, respecting the order in which variables take their values, and including variables changing their values over time as multiple variables (see also Gebharter and Kaiser 2014), we can make the temporal dimension explicit (see figure 2). Despite the possible increase in

---

[8] Analogous suggestions are made by Baumgartner and Gebharter (2016) and Baumgartner and Casini (forthcoming) to identify mechanistic constitution.

complexity, this adds quite some representational power to difference-making graphs while helping us use temporal order to distinguish different relevance relations.

In addition to temporal aspects, *scientific domain* can also give us useful clues as to what relations might mediate observed manipulability. The idea is perhaps best described by reference to the familiar levels-metaphor. Variables representing properties or events from the cognitive (mental, psychological) domain are usually regarded as on a "higher" level than, say, variables representing properties or events from the "lower"-level neurophysiological or genetic domains. My use of "levels" here is not tied to any specific account of levels; neither do I want to impose any specific hierarchy of domains. What matters for current purposes is merely whether two variables are *located in the same scientific domain*.[9] If they are not, we can draw on *systematic knowledge* (or assumptions we may have) about how variables from the domains in question relate and project that into our graphs.

For instance, cognitive processes are usually considered as neurophysiologically implemented by neural processing in the brain. Accordingly, variables representing mental processes, e.g. memory or depression, should be assumed to relate to variables representing, e.g., low serotonin or hippocampal long-term potentiation (LTP) by implementation rather than causation. Similarly, insights about *specific* containment relations (this AMPA receptor is located in the postsynaptic membrane of hippocampal CA1) can be used to distinguish difference-making mediated by causal dependencies from difference-making due to part-whole relations. Graphically, we can represent such insights by placing variables in our graphs along a vertical dimension and marking known relations with specific kinds of arrows (see figure 2).

In principle, heuristic inferences based on time and domain should be considered independent, neither is primary. Insights on time and domain are

---

[9] For current purposes, I am using the term "scientific domain" rather non-technically to refer to scientific fields or areas of research like neuroscience, psychology, astronomy, or genetics. However, scientific domains may cross-cut layers of a traditional layer cake picture of science and how to best individuate scientific domains may depend on the research project at hand (see Kästner 2018).

usually acquired in different ways: Information on time is typically gathered through intervention-based studies. Knowledge—or assumptions—about systematic relations between domains or specific containment relations, by contrast, is usually acquired through meta-scientific reasoning or theorizing as well as non-intervention studies. Non-intervention studies do not manipulate some factor X with respect to another factor Y but study features like structure and organization of a system by other means (see Kästner 2015). Examples include staining, tracing, cutting-open, centrifuging, and x-raying. Moreover, insights on time and domain tend to play different roles in our search for causal relations: The temporal order revealed by interventions tends to suggest candidate causal relations. Information about relations between concrete variables or between variables from different domains typically restricts which manipulability relations might be considered candidates for causal relations. At times, the two strategies may constrain one another or deliver conflicting results. When conflicting results occur, we must decide which one to prioritize based on the reliability of the information we fed into our heuristics to begin with.

Finally, we might take into account multiple realizers to acknowledge that, e.g., mental disorders may be neurophysiologically implemented in different ways. Depending on how a certain psychopathology is realized in a patient it may develop and be influenced in different ways. This is important to understand, say, why some patients respond well to certain treatments and others do not. Graphically, we can capture this in different planes along a third dimension of difference-making graphs where the structure of the graph may differ between planes.

Visualizing our insights from these different heuristics in difference-making graphs supplied with dimensions and multiple kinds of arrows helps us construct integrative network explanations in psychiatry without loosing sight of actual causal relations. Figure 2 illustrates what this may look like for Tom's case.
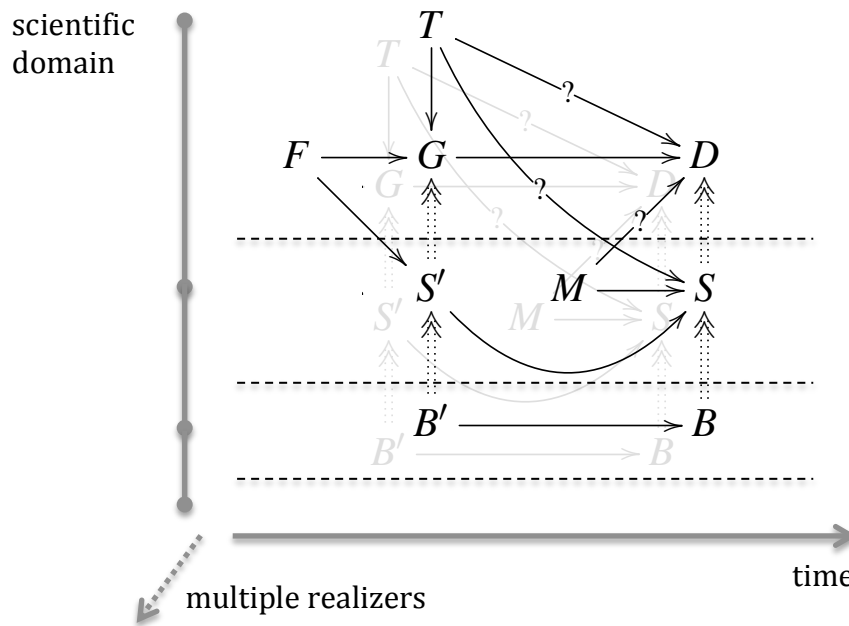
Figure 2: A difference-making graph of Tom's case. CBT (T) influences Tom's grief (G) and possibly his depressive state (D) as well as his serotonin levels (S). Gene expression as measured using biomarkers (B', B) may be considered the implementation base or realizer (dotted double arrows) of serotonin levels (S', S). Medication (M) targets serotonin levels but may also have a cognitive effect (possibly a placebo-effect) on D. Different kinds of relations are marked with different kinds of arrows and variables are arranged along three dimensions (time, domain, realizers). The grey graph symbolizes the inclusion of multiple realizers in different planes; causal relations may differ between these planes.

4. Conclusions

Causal explanations in psychiatry are not merely a result of "sloppy talk"; distinguishing between causes and other difference-makers is actually highly relevant for psychiatry as a scientific discipline as well as for clinical practice. However, multiple different factors may contribute to a given psychopathology in various ways. Thus restricting our analysis to causal relations *only* is too limited. DMI acknowledges this by modifying the Woodwardian interventionism such that it can be applied in contexts where non-causal dependencies are present. Still, identifying causal relations among other dependencies remains an epistemically demanding endeavor. Employing additional heuristics based on multiple experiments, considerations of time and scientific domain may help us identify genuine causes among difference-makers.

References

Baumgartner, M. (2010). Interventionism and epiphenomenalism. Canadian Journal of Philosophy, 40 (3), 359–383.

Baumgartner, M. (2013). Rendering interventionism and non-reductive physicalism compatible. Dialectica, 67, 1–27.

Baumgartner, M. & Casini, L. (forthcoming). An Abductive theory of constitution. Philosophy of Science.

Baumgartner M. & Gebharter, A. (2016). Constitutive relevance, mutual manipulability, and fat-handedness, British Journal for the Philosophy of Science, 67, 731–756.

Beck, A.T. & Alford, B.A. (2009). *Depression: Causes and Treatment*. Philadelphia, PA: University of Pennsylvania Press.

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, 16, 5–13.

Borsboom, D., Cramer, A. & Kalis, A. (forthcoming). Brain disorders? Not really... Why network structures block reductionism in psychopathology research. *Behavioral and Brain Sciences*.

Butler, A.C., Chapman, J.E., Forman, E.M. & Beck, A.T. (2006). The empirical status of cognitive-behavioral therapy: A review of meta-analyses. *Clinical Psychology Review*, 26, 17–31.

Campbell, J. (2007). An interventionist approach to causation in psychology. In: A. Gopnik & L. Schulz (eds.), Causal Learning: Psychology, Philosophy, and Computation. New York: Oxford University Press, pp. 58–66.

Campbell, J. (2016). Validity and the Causal Structure of a disorder. In: Kendler, K. and Parnas, J. (eds.) Philosophical Issues in Psychiatry IV: Psychiatric Nosology. Oxford: Oxford University Press.

Cuthbert, B.N., Lang, P.J., Strauss, C., Drobes, D., Patrick, C.J. & Bardley, M. (2003). The psychophysiology of anxiety disorder: Fear memory imagery. *Psychophysiology*, 40, 407-422.

Eronen, M. I. (2010). Reduction in Philosophy of Mind: a Pluralistic Account. Ph.D. thesis, University of Osnabrück.

Eronen, M. I. (2012). Pluralistic physicalism and the causal exclusion argument. European Journal for Philosophy of Science, 2, 219–232.

Gebharter, A. (2015). Causal exclusion and causal Bayes nets. Philosophy and Phenomenological Research. DOI:10.1111/phpr.12247.

Gebharter, A. & Kaiser, M. I. (2014). Causal graphs and biological mechanisms. In: M. I. Kaiser, O. Scholz, D. Plenge, & A. Hüttemann (eds.), *Explanation in the special sciences: The case of biology and history*, Dordrecht: Springer, pp. 55–86.

Hoffmann-Kolss, V. (2014). Interventionism and Higher-Level Causation. International Studies in the Philosophy of Science, 28, 49-64.

Kästner, L. & Andersen, L. (2018) Intervening into Mechanisms: Prospects and Challenges. *Philosophy Compass*. Manuscript under revision.

Kästner, L. (2015). Learning About Constitutive Relations. In: U. Mäki, S. Ruphy, G Schurz & I. Votsis (eds.), *Recent Developments in the Philosophy of Science: EPSA13 Helsinki,* pp. 155-167. Springer.

Kästner, L. (2017). *Philosophy of Cognitive Neuroscience: Causal Explanations, Mechanisms & Experimental Manipulations*. Berlin: Ontos/DeGruyter.

Kendler, K.S. & Campbell, J. (2009). Interventionist casual models in psychiatry: repositioning the mind-body problem. Psychological Medicine, 39, 881887.

Kim, J. (1998). Mind in a Physical World. Cambridge: MIT Press.

Lewis, D. (1970). Causation. Journal of Philosophy, 70, 556—567.

Pearl, J. (2000). Causality: Models, Reasoning, and Inference. Cambridge: Cambridge University Press.

Pies, R.W. (2014). The bereavement exclusion and DSM-5: An update and commentary. *Innovations in Clinical Neuroscience*, 11, 19–22.

Raatikainen, P. (2010). Causation, exclusion, and the special sciences. Erkenntnis, 73, 349–363.

Rescorla, M. (2017). An interventionist approach to psychological explanation. *Synthese*. DOI: 10.1007/s11229-017-1553-2

Shapiro, L.A. (2010). Lessons from causal exclusion. Philosophy and Phenomenological Research, 81, 594–604.

Shapiro, L. & Sober, E. (2007). Epiphenomenalism - The do's and the don'ts. In: P. Machamer & G. Wolters (eds.), Thinking about Causes. Pittsburgh: University of Pittsburgh Press. pp. 235–264.

Sofronoff, K. Attwood, T. & Hinton, S. (2005). A randomised controlled trial of a CBT intervention for anxiety in children with Asperger syndrome. *Journal of Child Psychology and Psychiatry*, 46, 1152-1160.

Spirtes, P., Glymour, C. & Scheines, R. (1993). Cauation, Prediction and Search. New York: Springer.

Wagner, B. (2014). *Komplizierte Trauer*. Berlin und Heidelberg: Springer.

Woodward, J. (2003). Making Things Happen. Oxford: Oxford University Press.

Woodward, J. (2008). Mental causation and neural mechanisms. In: J. Hohwy & J. Kallestrup (eds.), Being Reduced: New Essays on Reduction, Explanation, and Causation. Oxford: Oxford University Press, pp. 218–262.

Woodward, J. (2015). Interventionism and causal exclusion. Philosophy and Phenomenological Research, 91, 303-347.